acid repeat region of the A2AB gene and a 21-bp region of the *BRCA1* gene that is repeated up to four times. The aligned data sets were the following lengths: A2AB (1,164 bp); IRBP (1,292 bp); vWF (1,251 bp); 12S rRNA/tRNA valine-16S rRNA (2,001 bp); and BRCA1 (2,947 bp). Alignments for the concatenated (A2AB + IRBP + vWF + rRNA) and BRCA1 data sets are available in the Supplementary Information. Phylogenetic analyses included unweighted and transversion parsimony, minimum evolution (5,708-bp data set) or neighbour joining (BRCA1 data set) with logdet and maximum likelihood-GTR[24] distances, neighbour joining with weighted average (WAVE) maximum likelihood distances[25], and maximum likelihood under the HKY85 (ref. 24) model of sequence evolution. Gaps were coded as missing in parsimony analyses. Maximum-likelihood estimates of relative rates and transition to transversion ratios were obtained from maximum parsimony trees and used in subsequent maximum-likelihood analyses and in calculating weighted-average maximum-likelihood distances. Bootstrap support values are based on 500 replications except for maximum likelihood (100 replications). Maximum-likelihood bootstrap analyses with the BRCA1 data set used the following backbone constraint, where taxon numbers correspond to the ordering of taxa (top to bottom) in Fig. 1B: (((1–4), (5, 6), (7, 8)), ((9, 10), 11-14), (15, 16), (17, 18), 19–21, ((((22–24), 25), 26), 27), 28, (29, 30), 31–35, ((36, 37), (38, 39), (40, 41)), ((42, 43), 44), (45, 46), 47, ((48, 49), 50, 51)). Kishino and Hasegawa tests[24] were used to examine *a priori* hypotheses and to examine statistically acceptable root locations. In the latter case, we obtained the best unrooted likelihood tree for each data set and then evaluated all possible root positions. All phylogenetic analyses and statistical tests were performed with PAUP 4.0b2 (ref. 26), except for neighbour-joining with weighted average distances, where analyses were performed with PHYLIP 3.572 (J. Felsenstein) and WAVEBOOT (D. King and C. Krajewski). Maximum-likelihood analyses with rate partitions allowed the following eight rate partitions with the 5,708-bp data set: third positions of each nuclear gene; first + second positions of each nuclear gene; RNA stems; and RNA loops. Two rate partitions, corresponding to first + second and third codon positions, respectively, were used with the BRCA1 data set. NJ-WAVE analyses used a weighted-average distance approach[25] with the eight partitions indicated above for the 5,708-bp data set and two partitions (first + second codon positions; third codon positions) for the 2,947-bp data set; each partition was allowed its own rate, base composition, and transition to transversion ratio. Monte Carlo simulations were performed with Seq-Gen 1.1 (ref. 27) (Supplementary Information). Molecular dates were estimated using QDATE[28] (Supplementary Information).

Received 30 August; accepted 10 October 2000.

1. Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356,** 121–125 (1992).
2. Shoshani, J. & McKenna, M. C. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol. Phylogenet. Evol.* **9,** 572–584 (1998).
3. de Jong, W. W. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13,** 270–275 (1998).
4. Springer, M. S., Burk, A., Kavanagh, J. R., Waddell, V. G. & Stanhope, M. J. The interphotoreceptor retinoid binding protein gene in therian mammals: implications for higher level relationships and evidence for loss of function in the marsupial mole. *Proc. Natl Acad. Sci. USA* **94,** 13754–13759 (1997).
5. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388,** 61–64 (1997).
6. Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95,** 9967–9972 (1998).
7. Teeling, E. C. *et al.* Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* **403,** 188–192 (2000).
8. Asher, R. J. A morphological basis for assessing the phylogeny of the 'Tenrecoidea' (Mammalia, Lipotyphla). *Cladistics* **15,** 231–252 (1999).
9. Luckett, W. P. & Hartenberger, J.-L. Monophyly or polyphyly of the order Rodentia: Possible conflict between morphological and molecular interpretations. *J. Mammal. Evol.* **1,** 127–147 (1993).
10. Waddell, P. J., Cao, Y., Hauf, J. & Hasegawa, M. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-logdet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo and elephant. *Syst. Biol.* **48,** 31–53 (1999).
11. Penny, D., Masegawa, M., Waddell, P. J. & Hendy, M. D. Mammalian evolution: Timing and implications from using the logdeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* **48,** 76–93 (1999).
12. McKenna, M. C. & Bell, S. K. *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
13. Rainger, R. *Agenda for Antiquity: Henry Fairfield Osborn and Vertebrate Paleontology at the American Museum of Natural History, 1890–1935* (Univ. Alabama Press, Tuscaloosa, Alabama, 1991).
14. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
15. Foote, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. Jr Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283,** 1310–1314 (1999).
16. Rich, T. H. *et al.* A tribosphenic mammal from the Mesozoic of Australia. *Science* **278,** 1438–1442 (1997).
17. Mouchaty, S. K., Gullberg, A., Janke, A. & Arnason, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17,** 60–67 (2000).
18. Waddell, P. J., Okada, N. & Hasegawa, M. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48,** 1–5 (1999).
19. Rasmussen, D. T. in *The Evolution of Perissodactyls* (eds Prothero, D. R. & Schoch, R. M.) 57–78 (Oxford Univ. Press, Oxford, 1989).
20. Matthew, W. D. The Carnivora and Insectivora of the Bridger basin, middle Eocene. *Mem. Am. Nat. Hist.* **9,** 291–567 (1909).
21. Novacek, M. J. The skull of leptictid insectivorans and the higher-level classification of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **183,** 1–111 (1986).
22. Easteal, S. Molecular evidence for the early divergence of placental mammals. *BioEssays* **21,** 1052–1058 (1999).
23. Thompson, J. D., Higgins, G. D. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
24. Swofford, D. L., Olsen, G. P., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–492 (Sinauer, Sunderland, Massachusetts, 1996).
25. Krajewski, C., Blacket, M., Buckley, L. & Westerman, M. A multigene assessment of phylogenetic relationships within the dasyurid marsupial subfamily Sminthopsinae. *Mol. Phylogenet. Evol.* **8,** 236–248 (1997).
26. Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods) Version 4, (Sinauer, Sunderland, Massachusetts, 1998).
27. Rambaut, A. & Grassly, N. C. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13,** 303–306 (1997).
28. Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15,** 442–448 (1998).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank F. Catzeflis for tissue samples. This work was supported by the NSF (M.S.S.) and the TMR program of the European Commission (W.W.d.J.; M.J.S.).

Correspondence and requests for materials should be addressed to M.S.S. (e-mail: mark.springer@ucr.edu).

Molecular phylogenetics and the origins of placental mammals

William J. Murphy*†, Eduardo Eizirik*‡†, Warren E. Johnson*, Ya Ping Zhang§, Oliver A. Ryder‖ & Stephen J. O'Brien*

** Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland 21702, USA*
§ Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
‖ Center for Reproduction of Endangered Species, Zoological Society of San Diego, San Diego, California 92112, USA
‡ Department of Biology, University of Maryland, College Park, Maryland 20742, USA
† These authors contributed equally to this work

The precise hierarchy of ancient divergence events that led to the present assemblage of modern placental mammals has been an area of controversy among morphologists, palaeontologists and molecular evolutionists. Here we address the potential weaknesses of limited character and taxon sampling in a comprehensive molecular phylogenetic analysis of 64 species sampled across all extant orders of placental mammals. We examined sequence variation in 18 homologous gene segments (including nearly 10,000 base pairs) that were selected for maximal phylogenetic informativeness in resolving the hierarchy of early mammalian divergence. Phylogenetic analyses identify four primary superordinal clades: (I) Afrotheria (elephants, manatees, hyraxes, tenrecs, aardvark and elephant shrews); (II) Xenarthra (sloths, anteaters and armadillos); (III) Glires (rodents and lagomorphs), as a sister taxon to primates, flying lemurs and tree shrews; and (IV) the remaining orders of placental mammals (cetaceans, artiodactyls, perissodactyls, carnivores, pangolins, bats and core insectivores). Our results provide new insight into the pattern of the early placental mammal radiation.

The panoply of morphological, ecological and genomic diversity among extant mammals offers considerable potential for studies of speciation, adaptation, molecular evolution, genome organization and biogeography[1–3]. Studies on morphology and both mitochondrial and nuclear genes have revealed several higher level phylogenetic associations[2,4–7], but a full resolution of the earliest placental

divergences has yet to be accomplished. We obtained sequences from segments of 15 nuclear and three mitochondrial genes (9,779 base pairs (bp)) in 64 placental and two marsupial species (Table 1). Phylogenetic analyses of the concatenated data set using maximum parsimony, maximum likelihood and distance based (neighbour joining) methods all converged on a nearly identical, well supported topology defining four principal eutherian lineages (Fig. 1). The results affirm monophyly of traditional placental orders (except Artiodactyla and Insectivora), and also support some previously proposed, as well as new, superordinal clades.

Our analyses provide further support for the superordinal clade Afrotheria[4,5,8] (Fig. 1, clade I), and place it as the most basal placental divergence. Within Afrotheria, our data support a consistently strong clustering of Paenungulata[9](Hyracoidea, Sirenia and Proboscidea) and a nested Tethytheria[10] (Sirenia and Proboscidea). The next basal mammalian lineage in our trees was the Neotropical order Xenarthra (Fig. 1, clade II), whose monophyly was strongly supported in all analyses. Within Xenarthra, our data consistently united sloths (*Choloepus* spp.) and anteaters (*Myrmecophaga*, *Tamandua*) in the proposed suborder Pilosa[6], which was in turn the sister group to armadillos. Although all of our trees showed the placental root to be either between Afrotheria and all other placentals (neighbour joining and maximum likelihood), or within Afrotheria (maximum parsimony; on the branches leading to elephant shrews or the tenrec), we could not statistically reject a scenario with Afrotheria and Xenarthra as sister groups, separated from all other eutherians by a basal split ($P \geq 0.07$ in Kishino–Hasegawa (parsimony and likelihood) and Templeton tests). Furthermore, we could not statistically reject a scenario showing Xenarthra basal to all other placentals ($P \geq 0.50$ in Kishino–Hasegawa (parsimony and likelihood) and Templeton tests), suggesting that additional data collection will be required to resolve the placental root with a high level of confidence.

After these two basal divergences are two internal superordinal clades of eutherians (Fig. 1, clades III and IV), which were found as sister groups in all of our analyses. Clade III represents a novel group uniting rodents, lagomorphs, primates, tree shrews (Scandentia) and flying lemurs (Dermoptera). This is in contrast to molecular studies that show support for rodents being basal eutherians[7,11–13], which was possibly influenced by incomplete taxon sampling and extreme rate acceleration in this group. Although bootstrap support for this clade is not as high as for the other principal clades (neighbour joining = 93%, maximum likelihood = 85%, maximum parsimony = 64% and weighted parsimony = 73%), it is consistently resolved in all of our analyses. Our increased taxon sampling provides robust support for the monophyly of rodents, as well as their sister-group relationship with lagomorphs (cohort Glires[9,14]), a grouping that had been well established on morphological grounds[2] but has been contradicted or unresolved with molecular approaches[13,15,16].

The cohort Glires is, in turn, the sister group of primates, flying lemurs and tree shrews. Within primates, lemur (Strepsirhini) and tarsier (Tarsiiformes) were found to be sister taxa (bootstrap $\geq 80\%$) that were separated from anthropoids by a deep divergence. This contrasts with the widely held view that tarsiers are more closely related to anthropoids than to Strepsirhini[17]. Furthermore, although maximum parsimony and maximum likelihood analyses consistently supported the monophyly of Primates and the sister-group relationship between Dermoptera and Scandentia, distance-based trees suggested primate paraphyly by placing flying lemurs as the sister group to anthropoids, and by placing tree shrews basal among the three orders (Fig. 1). Kishino–Hasegawa (parsimony and likelihood) and Templeton tests failed to reject either of these hypotheses ($P > 0.1$ in all cases), which indicates that further sampling within Dermoptera and Scandentia may be required to fully resolve the deep, contemporaneous divergence among these three orders.

Seven remaining placental orders—cetaceans, artiodactyls, perissodactyls, carnivores, pangolins, bats, and core insectivores (hedgehogs, shrews and moles)—comprise a fourth superordinal lineage (Fig. 1, clade IV), a grouping that has received support from whole-mitochondrial genome sequence analysis (although no pangolin was sampled)[7]. Our results affirm the monophyly of Chiroptera (megabats plus microbats), which has been a topic of debate for over a decade[18]. These data also reject ($P < 0.0001$ for all tests) a direct relationship of bats with primates, flying lemurs and tree shrews, which were classically proposed to form the cohort Archonta[14]. Our data support the polyphyly of Insectivora[4,5,8], with shrews, moles and hedgehogs forming a monophyletic group provisionally termed 'Eulipotyphla'[19], whereas tenrecs clearly cluster within the superordinal clade Afrotheria. We find strong support for the traditional view of placing hedgehogs in a clade with shrews and moles (bootstrap $\geq 99\%$), which is in contrast to the basal position within eutherians observed with mitochondrial DNA (mtDNA) genomes[20].

The inter-ordinal relationships within clade IV were not fully resolved with high confidence, which is possibly due to the very rapid diversification of this group. However, our parsimony and maximum likelihood analyses agree with recent mtDNA data[7], suggesting a sister-group relationship between Chiroptera and Eulipotyphla at the base of this principal clade (see Supplementary Information). The pangolin (Pholidota) was the sister group to carnivores in combined (and many of the single-gene) analyses, a

Table 1 Characteristics of nuclear and mitochondrial gene segments

Gene	Nucleotides			Seq. div. range (%)	Ident. mouse vs human (%)	CI-MP	No. of species amplified	Amino acids		
	No. of bp	No. of var. sites	No. of PI					No. of res.	No. of var. sites	No. of PI
ADORA3	330	227	191	0.31–37.08	78.5	0.34	63	110	80	62
*ADRB2**	834	392	313	0.36–18.06	86.4	0.35	60	278	97	63
APP-3′UTR	690	454	299	0.32–31.16	90.5	0.51	63	–	–	–
ATP7A	690	468	375	0–41.66	87.1	0.41	64	230	167	129
BDNF	558	250	179	0.54–27.52	90.7	0.36	63	186	73	38
BMI1-3′UTR	340	175	78	0–16.78	93.4	0.65	56	–	–	–
CNR1	1,002	398	332	0.41–27.25	89.6	0.30	62	334	55	24
CREM-3′UTR	422	302	227	0–28.53	85.1	0.46	65	–	–	–
EDG1	978	458	365	0.10–19.26	88.5	0.35	58	326	98	62
PLCB4-3′UTR	337	290	249	0.33–70.40	78.9	0.44	65	–	–	–
*PNOC**	333	218	173	0–42.55	82.5	0.43	56	111	69	62
RAG1	774	353	303	0.78–23.98	87.1	0.32	56	258	66	47
*RAG2**	444	255	203	0.45–23.36	89.2	0.38	62	148	86	54
TYR	426	258	224	0–34.66	85.9	0.36	52	142	83	67
ZFX	204	79	62	0–22.99	93.6	0.41	57	68	14	5
mtDNA	1,417	929	731	0–31.69	81.7	0.25	66	–	–	–
Total	9,779	5,506	4,304					2,416	963	655

*No outgroup included in comparisons.
bp, base pairs; var., variable sites; PI, phylogenetically informative sites; res., amino-acid residues; Seq. div., percentage of nucleotide sequence divergence (Kimura 2-parameter); CI, consistency index of most parsimonious tree(s).

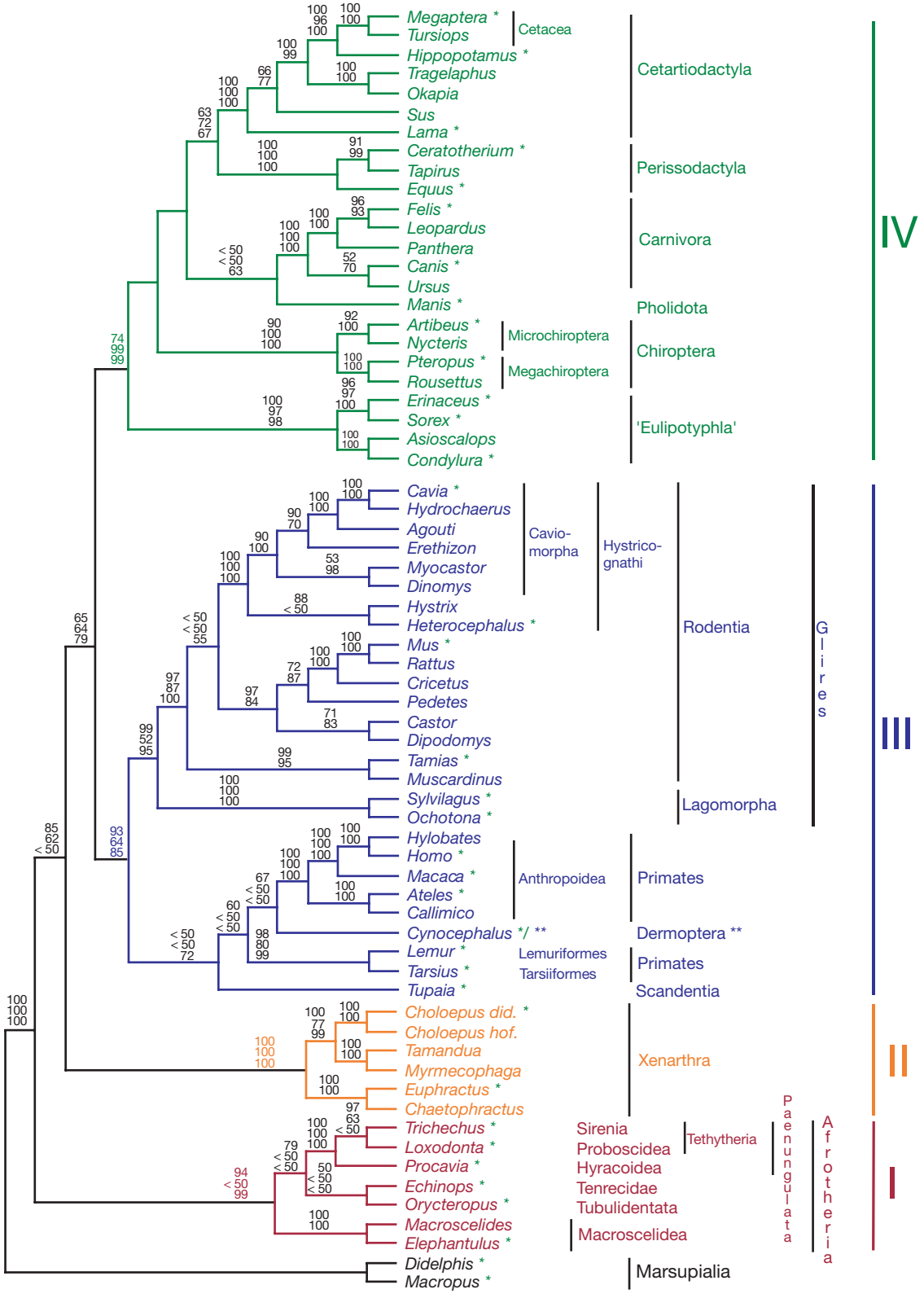

Figure 1 Phylogenetic relationships among 64 placental mammals and two marsupials based on analysis of 9,779 bp from 15 nuclear and three mtDNA genes. The tree represents the minimum evolution topology estimated through neighbour joining (NJ), using maximum likelihood (ML) distances (see Methods). Maximum parsimony (MP) and ML analyses (see Supplementary Information) produced a similar topology (MP: TL = 26,422, consistency index CI = 0.34, retention index RI = 0.47; ML: Ln Likelihood = –95086.03) with differences revolving mostly around short internodes with low bootstrap support. Numbers indicate per cent bootstrap support from NJ (top), MP (middle) and ML (bottom), based on 1,000 iterations for NJ and MP, and 100 iterations for ML. ML analyses were based on a pruned data set (37 taxa marked with asterisks). Terminal taxa are labelled by their generic designation, except where multiple members of a genus were included (see Methods). Brackets indicate higher level taxonomic groups observed in resulting trees (right of tree). A difference between MP/ML and NJ analyses was the position of the flying lemur (*Cynocephalus*, double asterisk). MP and ML differed from the shown tree by supporting primate monophyly and a sister-group relationship between Dermoptera and Scandentia. Weighted parsimony analyses using only transversions (Tv), removal of third position transitions (Ts), and a Tv:Ts weight of 2:1 produced topologies congruent with the shown tree, with differences revolving only around branches depicted here with low (< 50%) bootstrap support.

relationship that has also been suggested by previous morphological and molecular data[6,21,22]. Although our support for the Carnivora–Pholidota sister grouping fails statistical tests against alternative hypotheses (Kishino–Hasegawa (parsimony and likelihood) and Templeton tests; $P \geq 0.345$), our data do firmly place Pholidota within clade IV, and reject the view of an alliance between pangolins and the Neotropical Xenarthrans[2] (Kishino–Hasegawa (parsimony and likelihood) and Templeton tests, $P < 0.0017$). Our data also consistently indicated a sister-group relationship between Cetartiodactyla[19] and Perissodactyla, albeit with only moderate bootstrap support (63–72%).

We also performed phylogenetic analyses with the combined amino-acid sequences from the 11 coding nuclear genes, which provided medium to high support for the monophyly of most orders, as well as some superordinal relationships (for example, Afrotheria, Paenungulata, Glires, and a split between Afrotheria plus Xenarthra versus other placentals; see Supplementary Information). However, these data lacked sufficient power to resolve the other basal relationships identified with the nucleotide data set. This is not unexpected, given that the amino-acid data set contains less than 20% of the number of variable and parsimony-informative sites present in the nucleotide data set (Table 1).

Notably, our data set shows consistency of amplification and phylogenetic signal associated with the four non-coding nuclear segments (*APP*, *BMI1*, *CREM* and *PLCB4*) all of which were derived from 3′ untranslated regions (3′ UTRs). Whereas the average percentage of variable sites was expectedly higher for the 3′ UTRs (68% of 1,789 bp, $n = 4$) versus coding regions (51% of 6,573 bp, $n = 11$), the average consistency index was also higher for the 3′ UTRs (0.51) compared with the coding regions (0.33). Within the UTRs, we generally observed conserved blocks of sequence interrupted by variable (yet usually alignable) indels, which were often phylogenetically informative. It is probable that 3′ UTRs will be of importance in future phylogenetic studies of mammalian orders, as well as higher level phylogeny within other taxonomic groups.

The molecular resolution of placental mammals into four superordinal clades has been independently corroborated by an accompanying analysis[23]. Estimates for the timing of molecular divergence indicate that the superordinal diversification occurred 64–104 Myr ago (median 84 Myr ago[24]), several millions of years prior to the K/T boundary which marked the disappearance of the dinosaurs[25]. The earliest divergences (Afrotheria and Xenarthra) apparently occurred in the Southern Hemisphere, probably associated with the geological separation of the southern supercontinent Gondwanaland, followed by dispersal of the ancestors of Clades III and IV into the northern supercontinent of Laurasia[24]. The placement of several primitive insectivorous/generalist taxa near basal positions in each of the major clades (for example, Tenrecidae, Eulipotyphla and Scandentia; Fig. 1) is consistent with the early eutherian divergences preceding their remarkable morphological diversification. Geographical isolation of these primitive forms in different regions and the appearance of ecological niches vacated by the dinosaurs may account for the remarkable parallel adaptive radiations suggested by Madsen *et al.*[23]

From a genomics standpoint, these findings suggest that mouse and human, the two species at the forefront of genomic sequencing, are phylogenetically restricted to just one of the four principal placental lineages identified in our analyses. A broader knowledge of mammalian genomic organization, function and evolution will be gained by applying large-scale genome analysis to other species from each of these principal lineages, particularly Afrotheria and Xenarthra. This will achieve a better understanding of the history of our mammalian ancestors and the uniqueness of our own genome. □

Methods

Taxon sampling, amplification and sequencing

Our data set contains representatives from divergent lineages of all eutherian orders. Order Xenarthra: *Choloepus didactylus* (Linne's two-toed sloth), *Choloepus hoffmanni* (Hoffmann's two-toed sloth), *Tamandua tetradactyla* (tamandua), *Myrmecophaga tridactyla* (giant anteater), *Euphractus sexcinctus* (six-banded armadillo), *Chaetophractus villosus* (hairy armadillo); Order Insectivora: *Erinaceus concolor* (European hedgehog), *Sorex araneus* (European shrew), *Asioscalops altaica* (Old World mole), *Condylura cristata* (star-nose mole), *Echinops telfairi* (tenrec); Order Macroscelidea: *Elephantulus rufescens* (long-eared elephant shrew), *Macroscelides proboscideus* (short-eared elephant shrew); Order Tubulidentata: *Orycteropus afer* (aardvark); Order Hyracoidea: *Procavia capensis* (rock hyrax); Order Sirenia: *Trichechus manatus* (West Indian manatee); Order Proboscidea: *Loxodonta africana* (African elephant); Order Rodentia: *Tamias striatus* (eastern chipmunk), *Castor canadensis* (American beaver), *Muscardinus avellanarius* (dormouse), *Pedetes capensis* (springhare), *Mus musculus* (mouse), *Rattus norvegicus* (brown rat), *Cricetus griseus* (hamster), *Hystrix brachyurus* (Malayan porcupine), *Erethizon dorsatum* (North American porcupine), *Dipodomys heermanni* (kangaroo rat), *Heterocephalus glaber* (naked mole rat), *Cavia tschudii* (guinea-pig), *Hydrochaeris hydrochaeris* (capybara), *Myocastor coypus* (nutria), *Dinomys branickii* (pacarana), *Agouti taczanowskii* (mountain paca); Order Lagomorpha: *Ochotona hyperborea* (pika), *Sylvilagus cf. floridanus* (eastern cottontail); Order Dermoptera: *Cynocephalus variegatus* (Malayan flying lemur); Order Scandentia: *Tupaia minor* (lesser tree shrew); Order Primates: *Lemur catta* (ring-tailed lemur), *Tarsius bancanus* (Western tarsier), *Macaca mulatta* (rhesus macaque), *Ateles fusciceps* (brown-headed spider monkey), *Callimico goeldi* (Goeldi's monkey), *Hylobates concolor* (gibbon), *Homo sapiens* (human); Order Chiroptera: *Rousettus lanosus* (long-haired rousette), *Pteropus giganteus* (flying fox), *Nycteris thebaica* (slit-faced bat), *Artibeus jamacensis* (Jamaican fruit-eating bat); Order Carnivora: *Canis familiaris* (domestic dog), *Ursus arctos* (brown bear), *Felis catus* (domestic cat), *Leopardus pardalis* (ocelot), *Panthera onca* (jaguar); Order Perissodactyla: *Ceratotherium simum* (white rhinoceros), *Equus caballus* (horse), *Tapirus indicus* (Malayan tapir); Order Artiodactyla: *Lama glama* (llama), *Sus scrofa* (pig), *Hippopotamus amphibus* (hippopotamus), *Okapia johnstoni* (okapi), *Tragelaphus euryceros* (bongo); Order Cetacea: *Megaptera novaeangliae* (humpback whale), *Tursiops truncatus* (bottlenose dolphin); Order Pholidota: *Manis pentadactyla* (Chinese pangolin). For outgroup comparison we included both Australian (*Macropus eugenii*, Tammar wallaby) and American (*Didelphis virginianus*, opossum) marsupial representatives.

We designed primers as part of an ongoing effort to develop markers with uses in both mammalian comparative gene mapping and mammalian phylogenetics. The GenBank and UniGene databases (NCBI) were searched for genes with exons of sufficient length (> 200 bp) and variability (80–95% nucleotide identity between mouse and human), thereby providing adequate variation for the purpose of phylogenetic and somatic cell/radiation hybrid mapping. We designed primers in conserved stretches of the genes using Primer 3.0 (Whitehead Institute). Primer design avoided regions showing BLAST similarity to paralogues of the target gene. Primer pairs were selected by pre-screening each set in a panel of DNA from 11 individuals, including ten species from different eutherian Orders and one marsupial, using a touchdown PCR protocol with *Taq*-Gold DNA polymerase (PE Biosystems), and either 1.5 or 2.0 mM $MgCl_2$. Only those markers that showed amplification of a single, intense product of the predicted size in at least nine of the ten eutherian orders were chosen for the final study. We then chose successful primer pairs (see Supplementary Information) for amplification in a 96-well format (2 × 48 species or 6 × 16 species). We amplified remaining species in separate reactions with appropriate controls. Amplification products and Big Dye terminator (Applied Biosystems) sequencing reactions were purified using the *Psi*clone 96-well PCR purification system (Princeton Separations), and the 96-well Centri-sep system (Princeton Separations), respectively. Sequencing reactions were largely performed in a 96-well format using the ABI-3700 capillary sequencing apparatus, whereas smaller numbers of reactions were analysed using either an ABI-373 stretch or ABI-377 DNA sequencer.

Our initial data set consisted of > 12,000 bp from 20 nuclear-coding loci. We removed five genes (1,672 bp) owing to inconsistent amplification in > 20% of the species, failure of key taxa, or amplification of putative paralogous loci in some species. The final set of genes comprised 10,704 bp of which 925 bp were deleted because of ambiguous alignment, resulting in the total data set of 9,779 bp. A summary of the genes used and their characteristics is given in Table 1.

Data alignment and phylogenetic analysis

Phylogenetic analysis was performed on data sets aligned using the computer software CLUSTAL-X[26]. We manually inspected and modified alignment output. Regions of the alignments in which determination of homology was ambiguous were deleted before phylogenetic analyses. In the final data set, all taxa were represented by at least 11 of the 17 gene segments (see Supplementary Information). Phylogenetic analyses were performed using maximum parsimony, distance-based (neighbour joining), and maximum likelihood methods implemented in PAUP* 4.0b2 (ref. 27), PHYLIP 3.57, (DNAML and Protdist, J. Felsenstein, Univ. of Washington) and PUZZLE[28]. PAML 2.0 (Z. Yang, University College, London) was used for estimating the transition/transversion ratio (Ts/Tv = 2.0) and the α-parameter for rate variation among sites (α = 0.45). Parsimony analyses were performed using 50 replicates with random addition of taxa and tree-bisection reconnection branch swapping.

We performed minimum evolution (neighbour joining) analyses with several distance measures (Kimura 2-parameter, Logdet and maximum likelihood distances with HKY85 model) to examine effects on topological stability (see Supplementary Information). Maximum likelihood analyses for the nucleotide data set were performed with DNAML (assuming equal substitution rates among sites) and with PAUP* (using an HKY-γ model with estimated parameters); maximum likelihood trees for the amino-acid data set were constructed using PUZZLE (γ-corrected JTT model). Reliability of nodes was assessed using 1,000 bootstrap iterations for the nucleotide maximum parsimony and neighbour-

joining trees and the amino-acid maximum parsimony phylogenies, and 100 replicates for the nucleotide maximum likelihood tree and the amino-acid distance-based analyses (Dayhoff PAM matrix) (see Supplementary Information for additional trees and summary of bootstrap support). We performed tests of alternative phylogenetic hypotheses using Kishino–Hasegawa[29] (parsimony and likelihood) and Templeton's non-parametric[30] tests.

Received 30 October; accepted 4 December 2000.

1. Eisenberg, J. F. *The Mammalian Radiations* (Chicago Univ. Press, Chicago, 1981).
2. Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356,** 121–125 (1992).
3. O'Brien, S. J. *et al.* The promise of comparative genomics in mammals. *Science* **286,** 458–481 (1999).
4. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388,** 61–64 (1997).
5. Stanhope, M. J. *et al.* Highly congruent molecular support for a diverse clade of endemic African mammals. *Mol. Phylogenet. Evol.* **9,** 501–508 (1998).
6. McKenna, M. C. & Bell, S. K. *Classification of Mammals above the Species Level* (Columbia Univ. Press, New York, 1997).
7. Mouchatty, S. K., Gullberg, A., Janke, A. & Arnason, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17,** 60–67 (2000).
8. Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95,** 9967–9972 (1998).
9. Simpson, G. G. The principles of classification and a classification of mammals. *Bull. Am. Mus. Nat. Hist.* **85,** 1–350 (1945).
10. McKenna, M. C. in *Phylogeny of the Primates: a Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
11. Easteal, S. The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics* **124,** 165–173 (1990).
12. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
13. Reyes, A. *et al.* Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Mol. Biol. Evol.* **17,** 979–983 (2000).
14. Gregory, W. K. The orders of mammals. *Bull. Am. Mus. Nat Hist.* **27,** 1–524 (1910).
15. Graur, D., Duret, L. & Gouy, M. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). *Nature* **379,** 333–335 (1996).
16. Huchon, D., Catzeflis, F. M. & Douzery, E. J. P. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. *Mol. Biol. Evol.* **16,** 577–589 (1999).
17. Kay, R. F., Ross, C. & Williams, B. A. Anthropoid origins. *Science* **275,** 797–804 (1997).
18. Pettigrew, J. D. Flying Primates? Megabats have advanced pathway from eye to midbrain. *Science* **231,** 1304–1306 (1986).
19. Waddell, P. J., Okada, N. & Hasegawa, M. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48,** 1–5 (1999).
20. Krettek, A., Gullberg, A. & Arnason, U. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. *J. Mol. Evol.* **41,** 952–957 (1995).
21. Rose, K. D. & Emry, R. J. in *Mammal Phylogeny: Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 81–102 (Springer, New York, 1993).
22. Miyamoto, M. M. & Goodman, M. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. *Syst. Zool.* **35,** 230–240 (1986).
23. Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* (in the press).
24. Eizirik, E., Murphy, W. J. & O'Brien, S. J. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* (in the press).
25. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381,** 226–229 (1996).
26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25,** 4876–4882 (1997).
27. Swofford, D. L. *PAUP* Phylogenetic Analysis using Parsimony and Other Methods* (Sinauer, Sunderland, Massachusetts, 1998).
28. Strimmer, K. & Von Haeseler, A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13,** 964–969 (1996).
29. Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29,** 170–179 (1989).
30. Templeton, A. R. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37,** 221–244 (1983).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*. Sequences are deposited in GenBank under accession numbers AY011125–AY012154.

Acknowledgements

We thank D. Hirschmann, M. Houck, R. Montali, R. Baker, G. Harris, K. Helgen, A. L. Roca, M. Roelke-Parker, A. Grafodatsky, O. Serov and T. Oleksyk for help in obtaining samples and technical assistance, and M. Smith and M. Dean for helpful suggestions. We also thank the NCI Frederick Molecular Technology Center for technical support, and the Advanced Biomedical Computer Center for computational assistance. All tissue samples were obtained with appropriate permits (CITES) issued to the National Cancer Institute, National Institutes of Health (principal officer, S.J.O). Y.P.Z. is supported by the Natural Science Foundation of China and Chinese Academy of Sciences. E.E. is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

Correspondence and requests for materials should be addressed to S.J.O. (e-mail: obrien@mail.ncifcrf.gov).

Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants

Kathleen M. Pryer*, Harald Schneider*, Alan R. Smith†, Raymond Cranfill†, Paul G. Wolf‡, Jeffrey S. Hunt* & Sedonia D. Sipes‡

* *Department of Botany, The Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, Illinois 60605, USA*
† *University Herbarium, University of California, 1001 Valley Life Sciences Building 2465, Berkeley, California 94720, USA*
‡ *Department of Biology, 5305 Old Main Hill, Utah State University, Logan, Utah 84322, USA*

Most of the 470-million-year history of plants on land belongs to bryophytes, pteridophytes and gymnosperms, which eventually yielded to the ecological dominance by angiosperms 90 Myr ago[1–3]. Our knowledge of angiosperm phylogeny, particularly the branching order of the earliest lineages, has recently been increased by the concurrence of multigene sequence analyses[4–6]. However, reconstructing relationships for all the main lineages of vascular plants that diverged since the Devonian period has remained a challenge. Here we report phylogenetic analyses of combined data—from morphology and from four genes—for 35 representatives from all the main lineages of land plants. We show that there are three monophyletic groups of extant vascular plants: (1) lycophytes, (2) seed plants and (3) a clade including equisetophytes (horsetails), psilotophytes (whisk ferns) and all eusporangiate and leptosporangiate ferns. Our maximum-likelihood analysis shows unambiguously that horsetails and ferns together are the closest relatives to seed plants. This refutes the prevailing view that horsetails and ferns are transitional evolutionary grades between bryophytes and seed plants[7], and has important implications for our understanding of the development and evolution of plants[8].

Estimates of a phylogeny for the main groups of land plants, each with highly divergent morphologies, have been many, and all have been contested. Bryophytes (liverworts, hornworts and mosses) are consistently shown to be a basal grade, but their relationships to one another and to vascular plants are controversial[1,2,9–13]. Most phylogenetic analyses of vascular plants consistently reconstruct two main lines of evolution: the Lycophytina (clubmosses and relatives), with 1% of extant diversity, and the Euphyllophytina (all other vascular plants)[1,2,10,11,14–17]. Extant Euphyllophytina[1,2] comprises six major monophyletic lineages: Equisetopsida (horsetails), Polypodiidae (leptosporangiate ferns), Spermatophytata (seed plants), Psilotidae (whisk ferns; simple plants regarded by some to be living relicts of the earliest vascular plants[7,18]), Marattiidae and Ophioglossidae (eusporangiate ferns). Phylogenetic assessments based on single genes[10,11,14,15,19] and/or morphology[1,7,12,17,20] have provided only weak and usually contradictory evidence for the relationships among these euphyllophyte lineages. Resolving these relationships would not only stabilize a pivotal region of vascular plant phylogeny but is also key to identifying the most appropriate outgroup for addressing questions related to the evolution of seed plants.

Recent palaeontological studies[1,2,7] attempted to demonstrate that approaches based solely on living species would have difficulties reconstructing relationships among major lineages of vascular plants. Inadequate taxon sampling, rate heterogeneity across DNA nucleotide sites among lineages, and inappropriate algorithms also have been cited as impediments to resolving ancient branching events[21]. However, as predicted by recent theoretical studies[22], combined analysis of DNA sequences from multiple loci proves to